

## Use of Ordinal Categorical Variables in Skeletal Assessment of Sex From the Cranium

LYLE W. KONIGSBERG<sup>1\*</sup> AND SAMANTHA M. HENS<sup>2</sup>

<sup>1</sup>*Department of Anthropology, University of Tennessee,  
Knoxville, Tennessee 37996-0720*

<sup>2</sup>*Department of Cell Biology and Anatomy, Johns Hopkins University,  
School of Medicine, Baltimore, Maryland 21205*

**KEY WORDS**     probit analysis; visual sexing; Averbuch Site; sexual dimorphism

**ABSTRACT**     In anthropological studies, visual indicators of sex are traditionally scored on an ordinal categorical scale. Logistic and probit regression models are commonly used statistical tools for the analysis of ordinal categorical data. These models provide unbiased estimates of the posterior probabilities of sex conditional on observed indicators, but they do so only under certain conditions. We suggest a more general method for sexing using a multivariate cumulative probit model and examine both single indicator and multivariate indicator models on a sample of 138 crania from a Late Mississippian site in middle Tennessee. The crania were scored for five common sex indicators: superciliary arch form, chin form, size of mastoid process, shape of the supraorbital margin, and nuchal cresting. Independent assessment of sex for each individual is based on pubic indicators. The traditional logistic regressions are cumbersome because of limitations imposed by missing data. The logistic regression correctly classified 66/74 males and 46/64 females, with an overall correct classification of 81%. The cumulative probit model classified 64/74 males correctly and 51/64 females correctly for an overall correct classification rate of 83%. Finally, we apply parameters estimated from the logit and probit models to find posterior probabilities of sex assignment for 296 additional crania for which pubic indicators were absent or ambiguous. *Am J Phys Anthropol* 107:97–112, 1998.

© 1998 Wiley-Liss, Inc.

Sex determination in skeletal samples is a focal point of bioarchaeological research. In traditional life table analyses, anthropologists are concerned with age structure patterns by sex as well as biases that may occur in sex ratio estimates. In newer approaches from hazards analysis, the sex of the individual may be an important covariate to consider when modeling the hazard. Accurate determination of sex is also important in studies of morphological variation as well as in biocultural analyses that focus on differential sex roles and statuses. Indeed, there is little from human skeletal biology that does not rest on initial determinations of the sex of individuals.

Despite the importance of accurate and unbiased determinations of sex from osteological remains, the early history of sex assessment is largely based on anecdotal comments. For example, Hrdlička (1920:92) noted, "On the whole it may be said that an experienced and careful observer will have little if any difficulty in correctly identifying

---

Contract grant sponsor: National Science Foundation Systematic Anthropological Collections; contract grant number: SBR-9307693; Contract grant sponsor: National Park Service; contract grant number: CX 5000-9-5943.

\*Correspondence to: Lyle W. Konigsberg, Department of Anthropology, 250 South Stadium Hall, University of Tennessee, Knoxville, TN 37996-0720. E-mail: LyleK@utk.edu

Received 18 December 1997; accepted 10 June 1998.

over 80 per cent of the crania" with regard to sex. He felt that if the mandible was present the percentage correctly sexed would rise to 90%, and if the whole skeleton was present the percentage correctly sexed would exceed 96%. His statements 19 years later (Hrdlička, 1939) are nearly identical, although neither publication gives any empirical support for his stated percentages. In referring to the sexing of skulls, Comas (1960:436) suggested, "The probabilities of success depend on the number of coincidental features, and even under the best of conditions, (that is, with complete skulls) control experiments show a 10% error or the impossibility of diagnosis." It is not clear from this reference what the "control experiments" were. Later work by Krogman (1962) and Stewart (1979) was much more empirically based, using dissection-room known-sex individuals in blind studies. It is interesting to note that both Krogman and Stewart were aware of the essentially Bayesian nature of sex assessments on cadaver samples, where there is often a strong prior probability that a given individual is male. Krogman (1962:112) wrote:

Years ago when I was at Western Reserve University I sexed a sample of 750 adult skeletons (white and Negro, male and female) from the Todd Collection. Lots of 50 were done at a time. This was a dissecting-room population and records of age, sex, and race were available. In the 15 lots of 50 each I scored as follows: entire skeleton 100%; pelvis alone 95%; skull alone 92%; pelvis plus skull 98%; long bones alone 80%; long bones plus pelvis 98%. Now, these are biased results. In a medical school the ratio of male to female cadavera is about 15:1. Hence, in any case in which I was in doubt, I had a 15:1 chance of being correct if I said male. Therefore, it is likely that all estimates should be lowered 5–10%, depending on the relative completeness of the material to be sexed.

Similarly, Stewart (1979:88) recognized the skewed sex ratio as a problem, writing:

Back in 1943 when I spent a short time teaching anatomy at Washington University, St. Louis, I tested my ability to correctly sex the skull and lower jaw by inspecting those of 100 American blacks from the Terry Collection, equally divided between the two sexes. My score was only 77 percent correct (Stewart, 1948, p. 317). In thinking about this test later, I decided that it was a good thing I had not taken the skulls in the order in which they were catalogued (they had been selected by a

preparator), because, with females being greatly outnumbered in dissecting-room collections, I could have improved my score whenever in doubt simply by recording my judgment as "male."

We assume here that Stewart meant that he should have guessed male for the first 50 and female for the next 50 individuals if he had taken the material in order. Later work using discriminant function analysis was more explicitly Bayesian. In all cases of which we are aware, the authors assumed an uninformative prior (equal probabilities that individuals were male or female) for sex to derive estimators that could be justified either by a maximum likelihood or a Bayesian approach. The discriminant function analyses were argued (correctly) to be more objective than the visual methods that previously had been applied.

Although it should be clear even from this very brief review that many of the issues surrounding bias and accuracy of sex assessments have been amply discussed, we contend that there are still a number of unresolved issues. These are

1. how to classify individuals from archaeological samples where appropriate reference collections of known sex may not be available;
2. how to incorporate information from skeletons that are incomplete or poorly preserved and consequently provide incomplete data; and
3. how to use visual sexing methods in a justifiable quantitative fashion.

The last issue forms the central focus for this paper. There are numerous discussions in the literature of the use of discriminant function analysis in determining sex (Birkby, 1966; Giles, 1964; Giles and Elliot, 1963; Henke, 1977; Holman and Bennett, 1991; Meindl et al., 1985; Robling and Ubelaker, 1997; Steele, 1976) as well as more recent work on the use of finite mixture analysis (Dong, 1997; Josephson et al., 1996; McKeown, 1996; Pearson et al., 1992). In our discussion, we make some comparisons to discriminant function analysis. However, the quantitative use of visual sexing data is less developed than the metric methods, and as a

consequence we direct our attention to morphological assessments of sex in this paper. Our intent is not to present a specific method or implementation of a method but rather to carefully consider the methodology for sex assessment from morphological observations. As a consequence of this focus, we caution the reader against uncritically applying our specific methods without first addressing issues of interpopulation variation in sexual dimorphism as well as problems of observer repeatability.

### Dichotomous assignment of sex and the naive estimator

We depart here and in all following discussions from a common practice (e.g., Steele and Bramblett, 1988) of referring to assessment of sex as the determination of gender. *Gender* refers to the roles and behaviors associated with the sexes. As such, gender usually cannot be determined from skeletal remains, while sex sometimes can. In this section we consider the case where assignments are made on a dichotomous basis (i.e., individuals are assigned as male or female based on visual examination). We let  $p_1$  be the probability that a real female is classified as a female and similarly  $p_2$  be the probability that a real male is classified as a male. These probabilities can be obtained from blind studies of some reference sample of known sex and are presumed to generalize to a target sample (i.e., the level of dimorphism is the same for the reference and target samples, though the samples may differ in their sex ratio). On looking at any given skeleton, the osteologist must make a choice. The investigator may decide on the basis of the indicator or indicators that the skeleton should be assigned as a male or alternatively as a female. We will write the number of skeletons assigned as female as  $N_F$ , the number assigned as male as  $N_M$ , and the total number of skeletons as  $N$ . From these scorings, the unbiased estimator for the number of females actually represented at the site is

$$\frac{N_M - Np_2}{1 - p_1 - p_2}, \quad (1)$$

and the unbiased estimator for the proportion of females at the site is

$$\frac{\frac{N_F}{N} + p_2 - 1}{p_1 + p_2 - 1}. \quad (2)$$

McLachlan (1992:31–32) gives an algebraically identical form for equation 2, which he refers to as the “discriminant analysis estimator.”

We should contrast equation 2 with what could be called a naive estimator for the proportion ( $N_F/N$ ), where we assume no error in classification. This estimator is only unbiased if there is indeed perfect classification (i.e.,  $p_1 = p_2 = 1.0$ ) or if the actual sex ratio is 1:1 and classification is symmetric (i.e.,  $p_1 = p_2$ ). It may be reasonable to assume that most human sex ratios are near to 1:1, though there are archaeological samples that appear to deviate from this ratio because of mortuary practices or preservation. The presumption of equal classification rates for males and females is more problematic. Some indicators perfectly or nearly perfectly identify one of the sexes but not the other. For example, the presence of a so-called “groove of pregnancy” (Houghton, 1974) at the preauricular sulcus unambiguously identifies an individual as female, but the absence of such a groove identifies an individual as male or female. Similarly, the presence of a Y marker unambiguously (barring contamination of samples) identifies an individual as a male. However, absence of a Y marker could indicate either that the individual is a female or that he is a male and the marker went undetected. This latter possibility may be reduced to the level of being highly improbable by running additional extractions/amplifications (Stone et al., 1996). In the example discussed by Stone et al., where four PCRs are used if a Y chromosome is not detected on any one PCR, the naive estimator for the proportion of males would only be in error (too low) by about 6.7%. Though we can only hazard a guess here, we suspect that when sex is judged on the basis of rugosity,  $p_2$  is higher than  $p_1$  (i.e., the misclassification rate is higher for female skeletons than it is for male skeletons). As Weiss (1972) has noted,

this is the simplest explanation for the male-biased sex ratios he observed in the paleodemography literature. If we had reasonable guesses for the values of  $p_1$  and  $p_2$ , we could adjust sex ratios using equation 2. Unfortunately, any such information on misclassification rates is often highly anecdotal.

### Univariate seriation

The dichotomous assignment described above is quite limited in scope. With the exception of sex determination for isolated cases, we know of few examples where all sex determinations are unequivocally assigned on the basis of an indicator. Indeed, White (1991) and Meindl et al. (1985) have suggested that seriation instead be used to assign sex. If seriation is applied, then there must be some continuous but unspecified distribution to sex indicators. White (1991: 322) writes

When sexing skulls, always use the entire population under study. Seriate this population according to the criteria you use, and then sort.

However, such an approach is problematic. We can enumerate four specific problems with seriation. First, for a large series (the example we consider here has over 400 crania to be seriated), it is doubtful that crania could be seriated in any replicable manner. We do not know to what extent lack of repeatability would affect the final sex assignments. Second, if there are multiple cranial characters to consider, each must be examined in separate independent seriations, and it may not be possible to combine these into one seriation. Kendall (1966:171) noted this problem years ago:

It follows that any distribution-free method for dealing with the discrimination problem must rely on order properties. One of the distinguishing features of such properties, however, is that they exist only in one dimension.

Third, unlike the methods we will describe below, seriation does not provide posterior probabilities of sex (also referred to as probabilities of group membership in the discriminant literature) for individual cases. This is true because seriation is an inherently non-

parametric approach. Finally, because seriation ultimately assigns crania to sex as if all determinations were certain, standard errors on the proportions of sex will be underestimates (as in the naive method described above). The method appears to unambiguously assign sex when there may in fact be a considerable amount of ambiguity. While seriation may be a useful training/calibration method for working out ordinal scorings within skeletal collections (as Meindl et al. (1985) did), we question its practical applicability when actually attempting to sex large samples of human skeletal remains.

### Traditional logistic and probit methods for assigning sex on the basis of an ordinal categorical indicator

As we mentioned in the section on seriation, the dichotomous case (where an osteologist assigns an F or an M to a skeleton on the basis of some observation) is quite limited in its use. It is far more common for osteologists to use a three-point ordinal scale (F, ?, or M), a five-point scale (F, F?, ?, M?, and M), or a six-point scale (F+, F, F-, M-, M, and M+) (Meindl et al., 1985). The five-point scale has recently been recommended in *Standards for Data Collection from Human Skeletal Remains* (Buikstra and Ubelaker, 1994) for a large number of traits and earlier by the Workshop of European Anthropologists (1980). In this section, we focus on how a simpler single three-point ordinal scale trait might be used to assign sexes to individual skeletons and to estimate the sex ratio for a sample.

Walker et al. (1996) have suggested using logistic regression to determine sex from one or more ordinal categorical traits. We will take this suggestion as a point of departure. We presume that a single ordinal categorical trait has been scored on a three-point scale running from female to male for a series of known-sex individuals. We will use the symbol  $y_i$  to represent the scoring for the  $i$ th individual. This score can take the value 1 (F), 2 (?), or 3 (M). Given the observed score for an individual, we can find the probability



that the individual is a male by applying Bayes' Theorem:

$$p(m|y_i) = \frac{p(y_i|m) \times p(m)}{p(y_i|m) \times p(m) + p(y_i|f) \times p(f)}, \quad (3)$$

where  $p(f)$  is the prior probability that an individual is female,  $p(m)$  is the prior probability that an individual is male, and  $p(y_i|f)$  and  $p(y_i|m)$  are the probabilities of observing the indicator conditional on the individual being female or male, respectively. In the logistic regression model, it is assumed that the ratio of conditional probabilities is a simple exponential function, so

$$\frac{p(y_i|m)}{p(y_i|f)} = \exp(\alpha + \beta y_i). \quad (4)$$

If we solve equation 4 for  $p(y_i|f)$  and  $p(y_i|m)$ , substitute these into equation 3, and assume that  $p(f) = p(m) = 0.5$ , then we find in the logistic regression model that

$$p(m|y_i) = \frac{\exp(\alpha + \beta y_i)}{1 + \exp(\alpha + \beta y_i)}, \quad (5)$$

which is the usual result. In the closely related probit regression model, we replace the assumption shown as equation 4 (above) with

$$\frac{p(y_i|m)}{p(y_i|f)} = \frac{\Phi(\alpha + \beta y_i)}{1 - \Phi(\alpha + \beta y_i)}, \quad (6)$$

where  $\Phi(\alpha + \beta y_i)$  is a standard normal integral from  $\alpha + \beta y_i$  to infinity. Again, assuming that  $p(f) = p(m) = 0.5$ , we can find the posterior probability that an individual is a male as

$$p(m|y_i) = \Phi(\alpha + \beta y_i). \quad (7)$$

Now logistic regression (equation 5) or probit regression (equation 7) will give unbiased estimates of the posterior probability of sex conditional on an observed sex indicator if the equal prior ( $p(f) = p(m) = 0.5$ ) assumption is warranted and the form for the ratio of conditional densities (equation 4 for logistic or equation 6 for probit) is correct. In the next section, we suggest a method by which both of these assumptions can be removed in a cumulative probit model.

#### A cumulative probit model for assigning sex from a single ordinal categorical indicator

The cumulative probit and the closely related cumulative logit model are very commonly used tools for analyzing ordinal categorical data (Agresti, 1990; Aitchison and Silvey, 1957; Albert and Chib, 1993; Anderson, 1984; Arminger, 1995; Gurland et al., 1960; Long, 1997; McCullagh, 1980). The simplest way to think of the cumulative probit model in the sexing context is as an ordered multiple threshold model (this is illustrated in Fig. 1). Taking, for example, the scoring of brow ridges, we assume that the prominence of the brows is a continuous and normally distributed trait. Because we cannot directly form a quantitative measure of brow prominence, we instead score the brows by placing them into one of three ordered categories (F, ?, or M). On the latent scale (i.e., the unmeasured brow prominence expressed as a z score), a brow would be classified into one of the three categories depending on the position of its z score relative to thresholds dividing the categories. We represent the thresholds using the symbol  $\gamma$ . The separation between brows we would categorize as F vs ? is at  $\gamma_1$  and that between ? and M is at  $\gamma_2$ . Because the thresholds must be ordered, we have the constraint that  $\gamma_1 < \gamma_2$ . There is no threshold dividing F from any more feminine brow or M from any more masculine brow, so we can define  $\gamma_0$  and  $\gamma_3$  as negative infinity and infinity, respectively.

We now take the actual sex and code it with an indicator variable  $x$ , where male is coded as 0 and female as 1. The probability conditional on sex ( $x$ ) that an individual is in the  $j$ th category of the sex indicator (where  $j$  runs from 1–3) in the cumulative probit model is then

$$p(y_i = j|x_i) = \Phi(\gamma_{j-1} + \beta x_i) - \Phi(\gamma_j + \beta x_i). \quad (8)$$

Substituting equation 8 into equation 3 gives

$$p(m|y_i = j) = \frac{[\Phi(\gamma_{j-1}) - \Phi(\gamma_j)] \times p(m)}{[\Phi(\gamma_{j-1}) - \Phi(\gamma_j)] \times p(m) + [\Phi(\gamma_{j-1} + \beta) - \Phi(\gamma_j + \beta)] \times p(f)}. \quad (9)$$

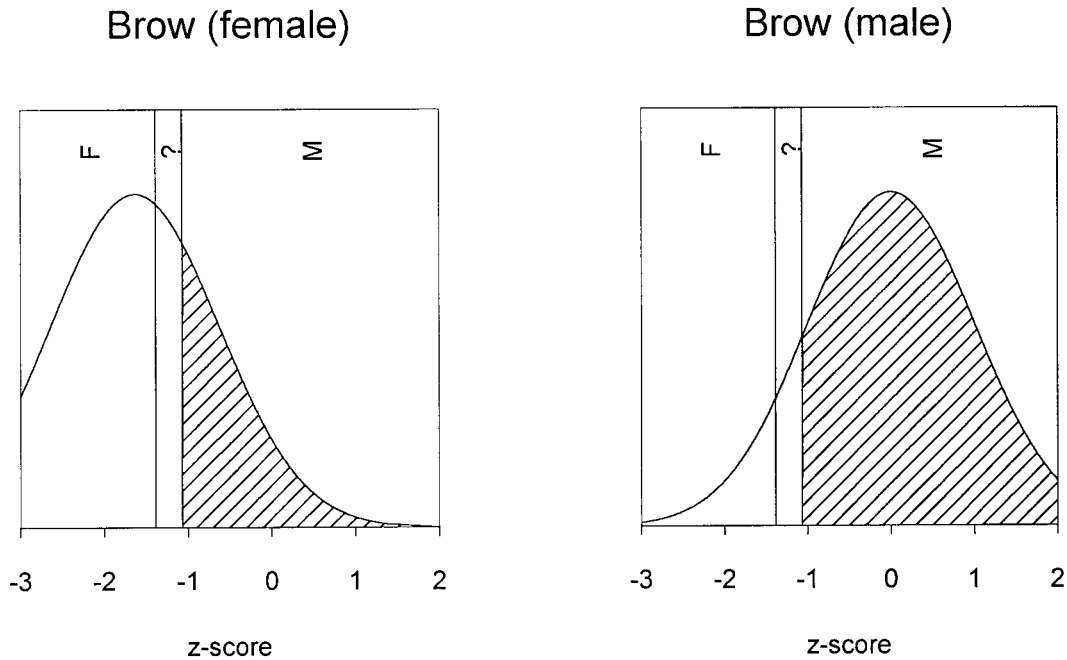


Fig. 1. Example of a cumulative probit analysis for prominence of the brow. There are two thresholds (one dividing F from ? and one dividing ? from M), and the male mean is constrained to equal 0.0. The shaded region represents the proportion of each sex expected to demonstrate a male brow.

In Materials and Methods, we discuss how equations 8 and 9 can be used within a maximum likelihood context to estimate posterior probabilities of sex and to estimate the proportional composition by sex of a sample. It is instructive at this point, however, to compare the cumulative probit model, as in equation 9, with the logistic regression model, as in equation 5. The logistic regression model assumes linearity in the logits. The cumulative probit model makes no such assumption of linearity, as the predictor variable (sex) is an observed binary variable. The cumulative probit model does assume that the thresholds for where an observer classifies a trait are fixed across sex, with only the means for males and females differing. This is a reasonable assumption and one that we can examine (as we do using a likelihood ratio test in Results).

#### Multivariate models for estimating sex from ordinal categorical indicators

The extensions of the traditional logistic or probit regressions to the multivariate

case are immediate. In equations 5 and 7, we would replace the simple regressions with multiple regressions on two or more indicators. These equations now require that the ratio of joint conditional probabilities follow multiple regression forms of equations 4 and 6. Walker et al. (1996) have already applied a multivariate logistic regression based on two sex indicators (glabella and mastoid form and nuchal and mastoid form). The multivariate extension for the cumulative probit is less simple. To fit such a model, equations 8 and 9 must be generalized to contain multivariate standard normal integrals. In Results, we present such models but only for the five-variable (five sex indicators) case. Because of the wide availability of high speed routines for numerically integrating multivariate normals (Drezner, 1992; Schervish, 1984) as well as the development of new Monte Carlo approaches to fitting multivariate probit regression models (Chib and Greenberg, 1997), we expect that higher dimensional problems may become commonplace in the future. One advantage of using the multivariate cumulative probit approach

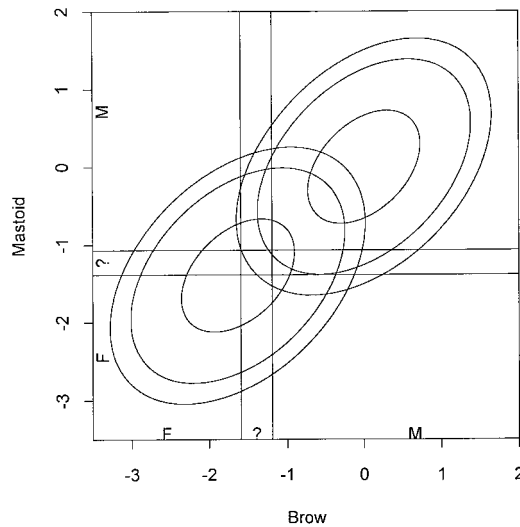


Fig. 2. Example of a bivariate cumulative probit analysis for prominence of the brow and mastoid process size. There are two thresholds (one dividing F from ? and one dividing ? from M) for each trait, and the male means are constrained to equal 0.0. The ellipses are the 25%, 50%, and 75% isodensities. The three ellipses to the lower left are for females, while the three ellipses to the upper right are for males.

vs. logistic or probit regression is that missing observations on indicators are easily accommodated. When an indicator cannot be observed, this is equivalent to saying that the integration across the unobserved indicator runs from negative to positive infinity. Consequently, the dimension of the integral is reduced by one for each indicator that is unobservable. Figure 2 shows an example of a bivariate cumulative probit model.

#### MATERIALS AND METHODS

During a routine inventory, all individuals from the Averbuch Site (40DV60), a Late Mississippian village located near Nashville, Tennessee, were assessed for morphological characteristics of the skull and pelvis commonly used in sex determinations (Bass, 1987; Buikstra and Ubelaker, 1994; Steele and Bramblett, 1988; White, 1991). Although there are approximately 880 individuals represented from this site (Berryman, 1981; Konigsberg and Frankenberg, 1994), many individuals could not be scored because they lacked both the skull and pelvis or because they were too young. As a

consequence, there were only 442 individuals for whom one or more cranial and/or pubic bone indicators of sex could be scored. Of these 442 individuals, 264 lacked sufficiently preserved pubic bones and are treated here as individuals of unknown sex. Of the remaining 178 individuals with pubic bones, 146 could be unambiguously assigned a sex. By *unambiguously* we mean that at least two of the Phenice (1969) characteristics were observable, and there was no discordance among the Phenice characteristics that could be observed. In this paper we concentrate on sexing from the skull, and consequently eight of the 146 individuals were eliminated from further study because they lacked observable crania. As the ages at death of the individuals are unknown, we have not included age as a covariate. Walker (1995) has demonstrated that age can have substantial effects on sexual dimorphism, with individuals appearing to be more masculine with increasing age.

For the remaining 138 individuals, five cranial variables are considered here. These variables are 1) superciliary arch form, 2) mastoid form, 3) superior orbital margin form, 4) nuchal cresting, and 5) chin form. These five variables correspond, respectively, to El-Najjar and McWilliams's (1978) prominence of the brow ridges, size of the mastoid process, orbital rim sharpness, muscular relief in the nuchal region, and shape of the chin. All traits were initially observed on a five point ordinal scale, of F, F?, ?, M?, and M, where these scorings can be loosely defined as unambiguously female, probably female, indeterminate, probably male, and unambiguously male, respectively. Both left and right sides were scored when present. To simplify this analysis, we have collapsed the scoring into a three point scale of female (F and F?), ambiguous (?), and male (M? and M). We have also randomly selected sides if both sides were observable. All observations were made by two observers (S.M.H. and Parvene Hamzavi) after initially standardizing by scoring a small set of skeletons in common. These two observers also undertook a large interobserver repeatability study which is reported elsewhere (Hens, 1994).

Although the crania and Phenice characteristics were scored at the same time, the observers had instructions to try to ignore information available from the os coxae when scoring the cranium (and vice versa).

We fit all logistic regression and probit models by numerically maximizing the log-likelihood using GEMINI (Lalouel, 1979). For the cumulative probit analysis, we found the multivariate normal integrals using a routine by Schervish (1984).

Although equation 3 can be applied to find the posterior probability that a single skeleton is male (or equivalently one minus the probability that it is female) conditional on observed indicators, it is not immediately apparent how to estimate the proportion of individuals who are a particular sex. If we substitute estimates for the proportion of individuals who are male ( $\hat{p}(m)$ ) and female ( $\hat{p}(f) = 1 - \hat{p}(m)$ ) into equation 3, then this allows us to find the posterior probabilities for all individuals. Because  $\hat{p}(m)$  is unknown, it is simplest to begin with the assumption that  $\hat{p}(m)$  is equal to 0.5 and then use equation 3 to find the posterior probabilities that skeletons are male. This is an expectation step of an EM algorithm (Dempster et al., 1977). The sum of the posterior probabilities divided by the total number of skeletons then gives a new estimate of  $\hat{p}(m)$ , and this process constitutes the maximization step. The new estimate of  $\hat{p}(m)$  can be inserted in equation 3 and the process repeated until convergence. To monitor convergence, the log-likelihood at any step can be found as (up to an additive constant)

$$\ln LK(\hat{p}(m)|y) = \sum_{i=1}^N \ln (\hat{p}(m) \times p(y_i|m) + (1 - \hat{p}(m)) \times p(y_i|f)), \quad (10)$$

and checked for change between EM cycles. At convergence, the posterior probability that any individual skeleton is male can be found from equation 3, and the log-likelihood from equation 10 can be compared to the initial log-likelihood (with  $\hat{p}(m)$  equal to 0.5) to test whether the sex ratio differs significantly from 1:1. Finally, the standard error of the estimate of the proportion of males (or equivalently of females) in the

sample can be found as

$$\text{s.e. } [\hat{p}(m)] = \text{s.e. } [\hat{p}(f)] = \left[ \sum_{i=1}^N \frac{[p(y_i|m) - p(y_i|f)]^2}{[\hat{p}(m) \times p(y_i|m) + (1 - \hat{p}(m)) \times p(y_i|f)]^2} \right]^{-1/2}. \quad (11)$$

Individuals of known sex (i.e., those sexed on the basis of the pubic bones) can be included in the standard error calculation by treating the conditional probabilities as equal to 0 or 1. For the logistic regression case, the conditional probabilities are not known, but one of the probabilities can be assumed equal to 1, while the other can then be obtained from equation 4.

For comparison to the logistic regressions and cumulative probit analyses, we also made very limited use of seriation. In particular, we seriated 93 of the 138 crania on the basis of the brow ridge. We also tried to seriate 109 of the 138 crania on the basis of the mastoid process but had to abandon this when we found that we were unable to order the crania on this feature. We did not attempt to seriate the crania on the basis of the orbital margins, nuchal crests, or mental eminence. For the brow ridge, we first broke the sample into three groups. The first group consisted of 19 crania that we felt were unambiguously female, the second group contained 13 crania that we felt were unambiguously male, and the remaining 61 crania were placed in a group between the female and male groups. We then subdivided the middle group into 25 possible females, 10 unidentifiable individuals, and 26 possible males. Following this division, we seriated within each of the groups, save for the 10 ambiguous crania, which we inserted into the final seriation of 83 crania. To find the optimal classification, we considered sectioning the sample at the first, second, third, and so on up to the ninety-third cranium.

As mentioned in the introduction, discriminant function analysis is a natural alternative to visual sexing. We calculated discriminant functions on crania in order to compare the functions' performance to visual sexing. Because these are archaeological crania that are often fragmentary and incomplete, there



TABLE 1. Logistic regression models for sex (posterior probability that an individual is male)<sup>1</sup>

Constant	Brow	Mastoid	Orbital	Nuchal	Chin	Sample size	Count in pattern	
							Sample 1	Sample 2
—	—	—	—	—	—	—	8	0
-2.5570	—	—	—	—	1.2320	129	2	29
-2.3988	—	—	—	1.1396	—	120	0	4
-4.3807	—	—	—	0.9321	1.1132	114	0	7
-2.0638	—	—	1.0764	—	—	116	0	2
-4.6971	—	—	1.0080	—	1.2753	110	0	2
-3.3967	—	—	0.9369	0.7550	—	107	0	0
-5.7393	—	—	0.9916	0.5326	1.2480	103	0	3
-3.4520	—	1.4312	—	—	—	132	0	13
-5.0574	—	1.1243	—	—	1.0766	125	6	25
-5.0509	—	1.2666	—	0.9205	—	119	1	9
-5.9680	—	0.9569	—	0.7700	0.9179	113	9	19
-4.5548	—	1.1088	0.9232	—	—	113	0	3
-6.2409	—	0.7200	0.9648	—	1.1884	108	0	5
-5.3954	—	1.1150	0.8720	0.5029	—	106	1	0
-6.5789	—	0.7457	0.9204	0.3635	1.0504	102	5	21
-3.3816	1.5123	—	—	—	—	114	1	4
-4.5824	1.2357	—	—	—	0.8457	107	0	1
-4.6017	1.4094	—	—	0.6645	—	104	0	0
-5.9467	1.1967	—	—	0.6585	0.8408	100	0	2
-3.9692	1.2259	—	0.6038	—	—	110	1	2
-5.4440	0.9222	—	0.6726	—	0.9431	105	1	8
-5.0307	1.2602	—	0.6644	0.4284	—	101	0	1
-6.7410	1.0094	—	0.7374	0.4244	0.9736	98	1	4
-5.1078	1.3337	0.8223	—	—	—	110	0	0
-5.8366	1.1673	0.5598	—	—	0.8238	105	0	5
-5.8086	1.1916	0.8358	—	0.5227	—	103	1	1
-6.4567	1.0375	0.5467	—	0.5684	0.7329	99	2	2
-5.4563	1.0991	0.7225	0.5628	—	—	107	1	4
-6.4946	0.8965	0.3953	0.6839	—	0.9612	103	6	26
-6.1379	1.0567	0.7882	0.6381	0.3023	—	100	3	14
-6.9643	0.8844	0.4226	0.6935	0.3534	0.8601	97	97	80

<sup>1</sup> Characters are scored as 1.0 (F), 2.0 (?), and 3.0 (M). Count in pattern is the number of crania in sample 1 (sexed from Phenice characteristics) and sample 2 (not sexed from the pubic bone) that fall into the pattern of observed/unobserved traits.

are substantial missing data problems with any craniometric approach. We have used a small set of six craniometrics (maximum cranial length, maximum cranial breadth, frontal breadth, bizygomatic breadth, basion-bregma height, and basion-nasion length), but even so 58 of the 138 crania have none of these measurements available. For the 80 remaining crania, we applied missing data discriminant analysis using a program written by Mark Schluchter (Little and Schluchter, 1985).

## RESULTS

### Traditional logistic regression

The traditional logistic regression approach leads to an uneconomical method for using and summarizing the observable cranial sexing information. To maximize the available information, we would want to use as many cranial variables as are observable. Because we also want to be able to sex

fragmentary remains, this requires that we produce logistic regressions for every possible missing data pattern. With five cranial variables, as in the current problem, there are 31 possible patterns ( $2^5 - 1$ , where one pattern is subtracted because it consists of no available observations). In the sample of 138 crania which can be sexed by associated os coxae, only about half (16 out of 31) of the possible missing data patterns are present. Table 1 gives the logistic regression parameters for the 31 possible regressions as well as the number of crania in the observed pattern (for both samples with and without sexable os coxae) and the number of crania on which the regression is based. The equations are easy to apply. For example, an individual who was scored as having a male brow and mastoid would (from equation 5) have a posterior probability of being male that is equal to  $0.7958 (= \exp(-5.1078 + 3 \times 1.3337 + 3 \times 0.8223) / (1 + \exp(-5.1078 +$

TABLE 2. Observed classifications (percents) and modeled percents<sup>1</sup>

Character	Sex	Count	Observed as			Cumulative probit			Prob.
			M	?	F	M	?	F	
Brow	M	60	85.00	6.67	8.33	85.37	6.09	8.53	0.8076
	F	54	29.63	11.11	59.26	29.28	11.71	59.01	
Mastoid	M	70	88.57	5.71	5.71	88.49	6.00	5.51	0.8829
	F	62	41.94	16.13	41.94	42.12	15.76	42.12	
Orbital margin	M	60	71.67	1.67	26.67	70.83	5.15	24.02	0.0586
	F	56	21.43	8.93	69.64	23.81	5.13	71.05	
Nuchal lines	M	66	77.27	12.12	10.61	78.95	8.07	12.99	0.0880
	F	54	37.04	7.41	55.56	34.31	12.44	53.25	
Chin	M	68	72.06	14.71	13.24	72.82	13.52	13.66	0.7176
	F	61	26.23	16.39	57.38	26.00	17.85	56.15	

<sup>1</sup> Prob. is the log-likelihood ratio goodness of fit test of the cumulative probit (three-parameter) model.

TABLE 3. Parameters from the multivariate cumulative probit model<sup>1</sup>

	F/? threshold	?/M threshold	Female mean	Pooled within-sex correlation matrix				
				Brow	Mastoid	Orbital margin	Nuchal lines	Chin
Brow	-1.3880	-1.0695	-1.6304	1.0000				
Mastoid	-1.5936	-1.1959	-1.3935	0.4136	1.0000			
Orbital margin	-0.7301	-0.5717	-1.2865	0.4756	0.2212	1.0000		
Nuchal lines	-1.1193	-0.8030	-1.1564	0.5357	0.3260	0.5426	1.0000	
Chin	-1.0911	-0.6050	-1.2725	0.5042	0.3377	0.0837	0.2600	1.0000

<sup>1</sup> The male mean is at 0 for all traits.

$3 \times 1.337 + 3 \times 0.8223$ ). The coefficients in this calculation are taken from Table 1, while the values of 3 represent the scores of male for the two traits. Reapplying the regressions in Table 1 to the 138 crania, 66 of the 74 male crania (89.19%) are sexed as male, while 46 of the 64 female crania (71.88%) are sexed as female. This gives an overall correct classification of 81.16%. This level of correct classification is probably slightly optimistic, because the training sample is being used as a test sample; however, with 138 crania, the influence of any one cranium on the functions should be small.

#### Cumulative probit model

Before applying the cumulative probit model, we should examine whether this model adequately fits the observed data. For any single character, such as the sharpness of the orbital margin, a four-parameter cumulative probit model will perfectly reproduce the observed data. In this model, there are two thresholds (dividing F from ? and ? from M) within each sex. The reduced model (equation 8, above) has three parameters, these being two thresholds (again, dividing F from ? and ? from M) and one parameter

that represents an offset of the mean for females. The likelihood ratio test for this reduced model against the full model consequently has one degree of freedom. Table 2 gives the asymptotic *P* values under a chi-square distribution as well as the observed and predicted percentages for scorings. The *P* values give a goodness of fit test for the cumulative probit model shown in equation 8. None of the five tests are significant at the 0.05 level, though the sharpness of the orbital margin and the prominence of the nuchal lines are both significant at the 0.1 level.

Table 3 gives the multivariate cumulative probit parameters fit using the 138 crania. These parameters were found using a numerical integration routine for the standard multivariate normal and a variable metric routine to search the likelihood surface. All of the correlations shown in Table 3 are positive, as would be expected if the cranial indications of sex are correlated beyond that predicted by actual sex. Table 4 gives the eigenstructure for the correlation matrix from Table 3. Only the first component shows an eigenvalue which is appreciably greater than 1.0, again demonstrating that there is

TABLE 4. Principal components (PC) of the correlation matrix from Table 3

	PC1	PC2	PC3	PC4	PC5
Brow	0.5364	0.0752	-0.2359	-0.7553	-0.2836
Mastoid	0.3957	0.3365	0.8507	0.0665	-0.0468
Orbital margin	0.4211	-0.5995	-0.0217	0.4407	-0.5182
Nuchal lines	0.4889	-0.3293	-0.0551	0.0292	0.8054
Chin	0.3733	0.6429	-0.4661	0.4795	-0.0129
Eigen-value	2.5176	1.0245	0.6735	0.4391	0.3453
% variance	50.35	20.49	13.47	8.78	6.91
cumulative %	50.35	70.84	84.31	93.09	100.00

considerable redundancy in the cranial indicators of sex. When the parameters from Table 3 are reapplied to the 138 crania, 64 of the 74 males (86.49%) are correctly classified as males, while 51 of the 64 females (79.69%) are correctly classified as females. This gives an overall correct classification of 83.33%, which is slightly better than the 81.16% observed from the logistic regressions.

#### Sex determination from univariate seriation of the brow ridges

Figure 3 show the cumulative counts against ranks from the seriation of brow ridges for females ( $n = 44$ ) and males ( $n = 49$ ). Here the females are shown as a cumulative distribution, while the males are shown as the complement of their cumulative distribution. As an example of how to interpret this figure, if we chose to section this sample at the tenth cranium (i.e., assign the first through tenth crania as female), then we would correctly assign nine of the 44 females as female (the third skull in the seriation is, in fact, a male). Conversely, we would correctly assign 48 of 49 male crania as male. Figure 3 also shows the combined correct number sexed at any given place in the seriation. There is no unique maximum from this figure. If we sectioned the sample at the forty-ninth, fifty-third, or fifty-seventh cranium, we would in each case get a correct percent classification of about 84%.

#### Sex determination from missing data discriminant analysis

Table 5 gives the missing data patterns for the six craniometrics on the 138 sexed crania. The discriminant function calculated on the 80 crania which had one or more measurements available resulted in an overall percent correct classification of 77.50%.

#### Sex determination for additional crania

The parameters from Tables 1 or 3 can be used to find the posterior probabilities of sex assignment for each of the 296 crania from Averbuch that could not be sexed from associated os coxae. This information can in turn be combined with the assigned sexes for the remaining 146 individuals that could be sexed from os coxae. Before examining the final sex ratios, it will be useful to look at the sex ratio that we would obtain using the naive estimator for sex. Unfortunately, there is no clear way to make this estimator operational. We have assumed, following most published accounts, that priority in sexing would be given to the size of the brow ridges first, then the size of the mastoids, then the prominence of the nuchal lines, then the form of the chin, and finally the sharpness of the orbital margin. Consequently, whichever of these indicators is observable and unambiguous (in order of priority) we take as the best indication of sex when making a naive estimate. When an indicator is observable but ambiguous, we move on to the first indicator which is unambiguous. By this method, we arrived at 171 males and 113 females represented in the 296 crania (12 crania could not be sexed by this method [see Table 6]). Combined with the 79 males and 67 females who could be sexed from os coxae, this gives 250 males to 180 females. The comparable figures from the logistic regressions in Table 1 are 149.82 males and 146.18 females for the 296 crania. Fractional individuals arise because the number of individuals are obtained from the sums of the posterior probabilities. Combined with the 79 males and 67 females sexed from os coxae, this gives 228.82 males and 213.18 females. Using the EM algorithm and the cumulative probit information from Table 3, the estimated number of

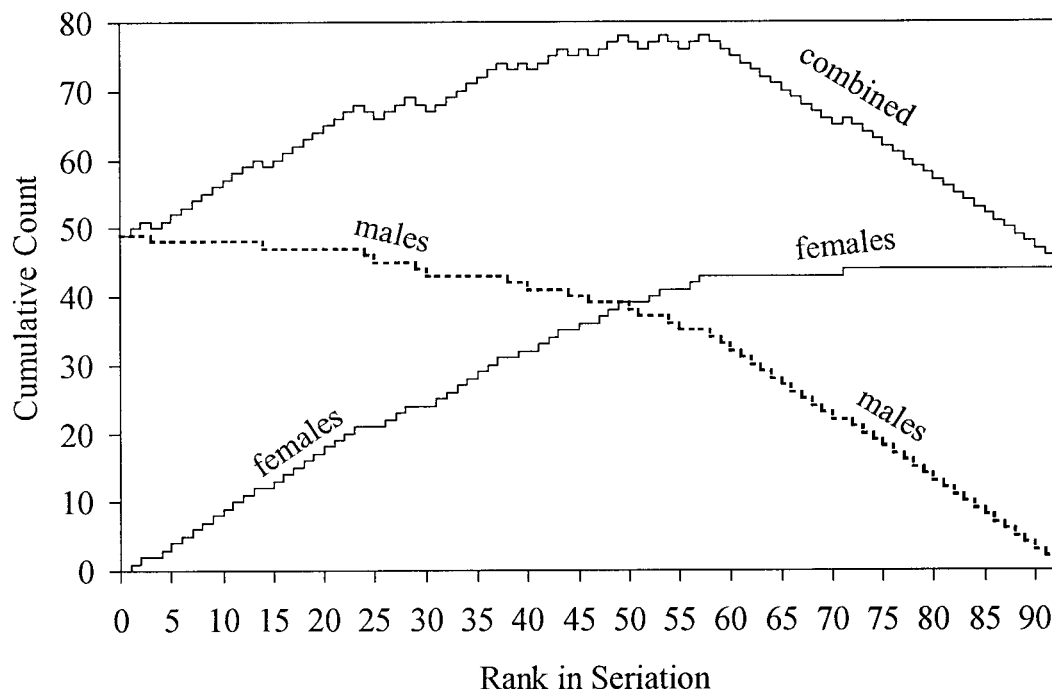


Fig. 3. Seriation results for 93 crania ranked on the basis of brow ridge size and morphology. There were 44 females and 49 males in this seriation. The lines for females (thin solid line) and males (thick dashed line) give the number of crania that would be correctly sexed for the sectioning point placed at each rank. The line (thin solid) labeled *combined* gives the total number of males and females correctly sexed at each rank used as a sectioning point.

TABLE 5. Missing data patterns for six craniometrics taken on a sample of 138 adult crania<sup>1</sup>

GOL	XCB	WFB	ZYB	BBH	BNL	Number in pattern
0	0	0	0	0	0	58
0	0	0	1	1	1	21
1	1	1	1	1	1	17
1	1	0	1	1	1	14
0	0	0	1	1	0	7
0	0	0	1	0	0	5
0	0	0	1	0	1	5
0	0	1	1	1	1	5
0	0	0	0	1	0	2
0	0	0	0	1	1	1
0	1	0	0	1	0	1
1	1	0	1	0	1	1
1	1	1	1	0	0	1

<sup>1</sup> A 0 in a column indicates that the measurement cannot be observed, while a 1 indicates that the measurement is observable.

males is 142.78, while the estimated number of females is 153.22. Combined with the individuals sexed from os coxae, this gives 221.78 males and 220.22 females. The counts from each of the three methods (naive, logistic regression, and cumulative probit) are

TABLE 6. Naive estimation of sex for the crania that could not be sexed from Phenice characteristics of the associated os coxae<sup>1</sup>

	Sexed as male	Sexed as female	Total
Brow	88	57	145
Mastoid	55	35	90
Nuchal lines	12	4	16
Chin	16	15	31
Orbital margin	0	2	2
Total	171	113	284

<sup>1</sup> Rows are listed in order of priority for sexing.

summarized in Table 7 in the form of estimated proportions of males. This table also gives the standard errors of these proportions and the approximate 95% confidence intervals.

## DISCUSSION

The results of this study do not inspire complete confidence in our ability to determine the sex of individuals from incomplete crania. At best, our percentage of correct

TABLE 7. Comparison of estimated proportion of males ( $p(m)$ ) from three estimators

Method	$p(m)$	s.e. ( $p(m)$ )	Lower 95%	Upper 95%
Naive estimator	0.5814	0.0238	0.5348	0.6280
Logistic regression	0.5177	0.0299	0.4591	0.5762
Cumulative probit	0.5018	0.0302	0.4426	0.5610

identifications for sex was only about 83% when we used the cumulative probit method. At worst, our percentage of correct classifications was about 79% from the naive estimator. This latter method is particularly troubling because it is very asymmetrical in its misclassifications, with females being much more frequently misclassified as males than vice versa. This pattern of asymmetry is opposite of that found by Meindl et al. (1985) but is in keeping with the asymmetry suggested in Weiss's (1972) results. Before further comment, we should warn the reader that our results should not be applied uncritically to other skeletal series. In order for our results to be generalized, it would be necessary that other series have the same level of sexual dimorphism as in the Averbuch collection and that observers score in a manner similar to our observations. We doubt that either situation would generally hold. In this sense, then, we present our work as a possible model for future analyses rather than suggest that the parameters in, say, Table 1 be used directly.

Some comments are in order here concerning the seriation results we obtained from the brow ridges. On the surface, these results are promising, as they yielded a correct percent classification of about 84%. We feel, however, that a fairer test of the method would be to return to the 93 crania, use one of the sectioning point crania (either the forty-ninth, fifty-third, or fifty-seventh cranium from the original seriation) as a reference, and then classify all crania as female that were deemed (independently) to be to the left of the reference and all crania as male that were deemed to be to the right. In any event, we should note that the seriation method is poorly conceived from a methodological standpoint and is likely impractical when the number of crania is much above 100. There are at least three different ways that one can form a seriation, and we do not

know how they compare. We chose to seriate using a divisive method, but it is also possible to seriate using an agglomerative method. Here one starts by ordering two crania and then continues to insert crania into the sequence. A third and more taxing method is to compare all pairs of crania and then form the ranking from the preference matrix (see Cook and Kress, 1992).

To put our results in broader perspective, Table 8 lists results from other studies that have used visual sexing methods. The classification rates in Table 8 are listed separately for males and females. The table also gives the overall and average classification rates, where the overall rate is the proportion of correctly sexed individuals out of the total, while the average rate is the rate that would pertain if the sex ratio of the sample was 50:50 (i.e., it is literally the average of the correct classification rates for males and for females). We also give the lower and upper 95% confidence interval points for the odds ratios of male correct classification to female correct classification. This odds ratio equals 1.0 if there is complete symmetry of classification (i.e., the male and female percent correct classifications are equal) and is greater than 1.0 when there is less misclassification of males than of females. Intervals which do not include 1.0 represent significant asymmetry in the classifications. The intervals were found using methods described in Agresti (1990).

It is not clear that comparisons between the studies summarized in Table 8 are entirely valid, because some characters are inherently more dimorphic and because we would expect the levels of sexual dimorphism to vary between samples. Samples which are from populations that are more dimorphic should, of course, give better classification rates. In any event, the overall percent correct classifications from Table 8 range from a high of 99% correct in the Loth and Henneberg (1996) study of mandibular ramal flexure down to a low of about 69% in the Donnelly et al. (in press) study. We would have to question the high value given by Loth and Henneberg (1996), as it exceeds the value for indicators directly related to



TABLE 8. Summary comparison of sexing methods<sup>1</sup>

Source	Sexed from	Number of males	Percent correct (males)	Number of females	Percent correct (females)	Overall percent correct	Average percent correct	Lower 95% odds ratio	Upper 95% odds ratio
Loth and Henneberg (1996)	Adult mandible (ramal flexure)	116	99.13	84	98.81	99.00	98.97	0.97	1.03
Meindl et al. (1985)	Adult skull and pelvis	59	94.92	41	100.00	97.00	97.46	0.90	1.01
Phenice (1969)	Adult pubis	180	95.56	95	96.84	96.00	96.20	0.94	1.04
Meindl et al. (1985)	Adult pelvis	59	93.22	41	100.00	96.00	96.61	0.87	1.00
Meindl et al. (1985)	Adult skull	59	89.83	41	95.12	92.00	92.48	0.85	1.05
Houghton (1974)	Adult preauricular sulcus	54	100.00	65	70.77	84.03	85.38	1.21	1.64
Current study, cumulative probit	Adult skull	74	86.49	64	79.69	83.34	83.09	0.93	1.26
Current study, naive estimator	Adult skull	74	91.89	64	64.06	78.98	77.98	1.18	1.73
Current study, discriminant	Adult cranium	43	81.39	37	72.97	77.50	77.18	0.88	1.41
Mittler and Sheridan (1992)	Fetal ilium auricular elevation	34	85.29	24	58.33	74.13	71.81	1.01	2.06
Donnelly et al. (in press)	Adult mandible (ramal flexure)	56	82.14	40	50.00	68.75	66.07	1.17	2.26

<sup>1</sup> Average percent correct is the simple average of male and female percents correctly classified. Confidence intervals for odds ratios give information on the asymmetry of misclassifications. See text for further description.

the functional anatomy of the reproductive organs and/or the ability to bear offspring (Houghton, 1974; Meindl et al., 1985; Phenice, 1969). Koski (1996) has already called into question Loth and Henneberg's (1996) methods, and the recent study by Donnelly et al. (in press) applying Loth and Henneberg's (1996) method failed to reproduce the high percent correct classification stated by Loth and Henneberg. If the Loth and Henneberg method (whether applied by them or by Donnelly et al.) is discarded, then the lowest overall correct classification is from Mittler and Sheridan's (1992) study of fetal/neonatal iliac auricular surfaces.

The 83% correct classification that we obtained using the cumulative probit method compares favorably to other studies listed in Table 8. While our percent correct classification is lower than for any of the studies based on adult skeletal material (save for the Donnelly et al. (in press) study), it is important to recall that we applied our method to an archaeological sample where many skulls were incomplete (see Table 1 for an enumeration of the missing data patterns). This study therefore stands in stark

contrast to other studies, which have been based on complete dissection-room specimens. Indeed, we would be surprised if morphological sexing based on incomplete archaeological or paleoanthropological skulls would ever achieve an accuracy much greater than about 80%. Rates higher than about 85% would place morphological sexing accuracy in the same range as discriminant function analysis from relatively complete crania (see Table 1 from Henke [1977]).

The discriminant analysis gave classification results (77.50% correctly sexed) that were slightly worse than naive visual sexing (78.98% correct). Given the fragmentary and incomplete nature of this collection, it is difficult to recommend traditional craniometric approaches for sexing.

Aside from issues of overall classification, our results show that naive methods for determining sex can lead to naive interpretations. There is marked asymmetry in classification in our results from traditional visual sexing, with females being misclassified much more frequently than males. This is a potential area where initial seriation of at

least a portion of the collection might have led to standards for scoring which were more symmetric. If the asymmetry we observed were carried over into a similar sample where 50% of the individuals were female, based on the results from Table 8 we would end up estimating that 36.08% of the individuals were female. Using cumulative probit analysis, this bias is all but eliminated, with the estimated percentage of females rising to 46.37%. It is interesting to note that of the methods that have significant asymmetry in classification, all misclassify females as males more frequently than the converse. Combining this problem of asymmetry in classification with the fact that traditional methods of visual sexing do not allow observers to quantify uncertainty in their sex assessments, and thus give a greatly deflated standard error around proportions of the sexes, we see little reason to perpetuate sex determination outside of parametric models. Such models for ordinal categorical data are admittedly complicated, but their benefits far outweigh the additional computational burden.

#### ACKNOWLEDGMENTS

We thank Ms. Parvene Hamzavi for help on this project. We give credit where credit is due to StatLib (<http://temper.stat.cmu.edu/>) for a listing of Mark Schervish's algorithm for the multivariate normal integral. The data reported on here were collected while conducting an inventory under the auspices of a National Science Foundation Systematic Anthropological Collections grant (SBR-9307693). We are grateful to Dr. John Yellen, Program Director for Archaeology at NSF, for his help at numerous junctures and for his appreciation of the long-term research value of osteological collections from prehistoric contexts. The Averbuch Site was excavated under National Park Service funding (contract CX 5000-9-5943) to the University of Tennessee, with Drs. Walter E. Klippel and William M. Bass as coprincipal investigators. We thank Elaine Hens and Dr. Richard L. Jantz for their helpful comments on an earlier draft of this paper.

#### LITERATURE CITED

- Agresti A (1990) *Categorical Data Analysis*. New York: John Wiley and Sons.
- Aitchison J, and Silvey SD (1957) The generalization of probit analysis to the case of multiple responses. *Biometrika* 44:131–140.
- Albert JH, and Chib S (1993) Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Assoc.* 88:669–679.
- Anderson JA (1984) Regression and ordered categorical variables. *J. R. Statist. Soc. B* 46:1–30.
- Arminger G (1995) Specification and estimation of mean structures: Regression models. In G Arminger, CC Clogg, and ME Sobel (eds.): *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum Press, pp. 77–183.
- Bass WM (1987) *Human Osteology: A Laboratory and Field Manual*, 3rd ed. Columbia, MO: Missouri Archaeological Society.
- Berryman HE (1981) *The Averbuch Skeletal Series: A Study of Biological and Social Stress at a Late Mississippian Period Site from Middle Tennessee*. Ph.D. dissertation, University of Tennessee, Knoxville.
- Birkby WH (1966) An evaluation of race and sex identification from cranial measurements. *Am. J. Phys. Anthropol.* 24:21–28.
- Buikstra JE, and Ubelaker DH (1994) *Standards for Data Collection from Human Skeletal Remains*. Fayetteville: Arkansas Archaeological Survey.
- Chib S, and Greenberg E (1997) *Analysis of Multivariate Probit Models*. St. Louis: John M. Olin School of Business, Washington University (tech report).
- Comas J (1960) *Manual of Physical Anthropology*. Springfield, IL: Charles C. Thomas.
- Cook WD, and Kress M (1992) *Ordinal Information and Preference Structures*. Englewood Cliffs, NJ: Prentice Hall.
- Dempster AP, Laird NM, and Rubin DM (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B* 39:1–38.
- Dong, Z. (1997) Mixture analysis and its preliminary application in archaeology. *J. Archaeol. Sci.* 24:141–161.
- Donnelly SM, Hens SM, Rogers NL, and Schneider KL (in press) A blind test of mandibular ramus flexure as a morphologic indicator of sexual dimorphism in the human skeleton. *Am. J. Phys. Anthropol.*
- Drezner Z (1992) Computation of the multivariate normal integral. *Trans. Math. Software* 18:470–480.
- El-Najjar MY, and McWilliams KR (1978) *Forensic Anthropology*. Springfield, IL: Charles C. Thomas.
- Giles E (1964) Sex determination by discriminant function analysis of the mandible. *Am. J. Phys. Anthropol.* 22:129–136.
- Giles E, and Elliot O (1963) Sex determination by discriminant function analysis of crania. *Am. J. Phys. Anthropol.* 21:53–68.
- Gurland J, Lee I, and Dahm PA (1960) Polychotomous quantal response in biological assay. *Biometrics* 16:382–398.
- Henke W (1977) On the method of discriminant function analysis for sex determination of the skull. *J. Hum. Evol.* 6:95–100.
- Hens SM (1994) *The Use of Ordinal Categorical Variables in Skeletal Assessment of Sex*. M.A. thesis, University of Tennessee, Knoxville.
- Holman DJ, and Bennett KA (1991) Determination of sex from arm bone measurements. *Am. J. Phys. Anthropol.* 84:421–426.
- Houghton P (1974) The relationship of the pre-auricular groove of the ilium to pregnancy. *Am. J. Phys. Anthropol.* 41:381–390.

- Hrdlička A (1920) *Anthropometry*. Philadelphia: Wistar Institute.
- Hrdlička A (1939) *Practical Anthropometry*. Philadelphia: Wistar Institute.
- Josephson SC, Juell KE, and Rogers AR (1996) Estimating sexual dimorphism by method-of-moments. *Am. J. Phys. Anthropol.* 100:191–206.
- Kendall MG (1966) Discrimination and classification. In PR Krsihnaiah: *Multivariate Analysis*. New York: Academic Press, pp. 165–185.
- Konigsberg LW, and Frankenberg SR (1994) Paleodemography: "Not quite dead." *Evol. Anthropol.* 3:92–105.
- Koski K (1996) Mandibular ramus flexure—indicator of sexual dimorphism? *Am. J. Phys. Anthropol.* 101:545–546.
- Krogman WM (1962) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Charles C. Thomas.
- Lalouel J-M (1979) GEMINI—A Computer Program for Optimization of General Nonlinear Functions. Salt Lake City: University of Utah, Department of Human Genetics, (technical report 14).
- Little RJA, and Schluchter MD (1985) Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika* 72:497–512.
- Long JS (1997) *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.
- Loth SR, and Henneberg M (1996) Mandibular ramus flexure: A new morphologic indicator of sexual dimorphism in the human skeleton. *Am. J. Phys. Anthropol.* 99:473–485.
- McCullagh P (1980) Regression models for ordinal data. *J. R. Statist. Soc. B* 42:109–142.
- McKeown AH (1996) *Craniometric Sex Estimation Using a Finite Mixture Missing Data Model*. M.A. thesis, University of Tennessee, Knoxville.
- McLachlan GJ (1992) *Discriminant Analysis and Statistical Pattern Recognition*. New York: John Wiley and Sons.
- Meindl RS, Lovejoy CO, Mensforth RP, and Don Carlos L (1985) Accuracy and direction of error in sexing of the skeleton: Implications for paleodemography. *Am. J. Phys. Anthropol.* 68:79–85.
- Mittler DM, and Sheridan SG (1992) Sex determination in subadults using auricular surface morphology: A forensic science perspective. *J. For. Sci.* 37:1068–1075.
- Pearson JD, Morrell CH, and Brant LJ (1992) Mixture models for investigating complex distributions. *J. Quant. Anthropol.* 3:325–346.
- Phenice TW (1969) A newly developed visual method of sexing the os pubis. *Am. J. Phys. Anthropol.* 30:297–302.
- Robling AG, and Ubelaker DH (1997) Sex estimation from metatarsals. *J. For. Sci.* 42:1062–1069.
- Schervish MJ (1984) Multivariate normal probabilities with error bound. *Appl. Statist.* 33:81–94.
- Steele DG (1976) The estimation of sex on the basis of the talus and calcaneus. *Am. J. Phys. Anthropol.* 45:581–588.
- Steele DG, and Bramblett CA (1988) *The Anatomy and Biology of the Human Skeleton*. College Station: Texas A&M University Press.
- Stewart TD (1979) *Essentials of Forensic Anthropology*. Springfield, IL: Charles C. Thomas.
- Stone AC, Milner GR, Pääbo S, and Stoneking M (1996) Sex determination of ancient human skeletons using DNA. *Am. J. Phys. Anthropol.* 99:231–238.
- Walker PL (1995) Problems of preservation and sexism in sexing: Some lessons from historical collections for paleodemographers. In SR Saunders and A Herring (eds.): *Grave Reflections: Portraying the Past Through Cemetery Studies*. Toronto: Canadian Scholar's Press, pp. 31–45.
- Walker PL, Drayer FJ, and Siefkin SK (1996) *Malibu Human Skeletal Remains: A Bioarchaeological Analysis*. Santa Barbara: Department of Anthropology, University of California—Santa Barbara.
- Weiss KM (1972) On the systematic bias in skeletal sexing. *Am. J. Phys. Anthropol.* 37:239–250.
- White TD (1991) *Human Osteology*. New York: Academic Press.
- Workshop of European Anthropologists (1980) Recommendations for age and sex diagnoses of skeletons. *J. Hum. Evol.* 9:517–549.